

LA-UR-21-20682

Approved for public release; distribution is unlimited.

Title: Titan on the Red (advanced AI / ML system)

Author(s): Ali, Alee Rizwan

Intended for: presentation to the NNSA SAGE group

Issued: 2021-01-26

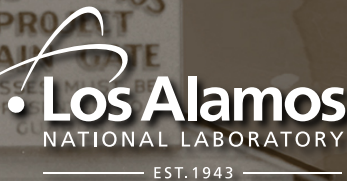
Disclaimer:

Los Alamos National Laboratory, an affirmative action/equal opportunity employer, is operated by Triad National Security, LLC for the National Nuclear Security Administration of U.S. Department of Energy under contract 89233218CNA000001. By approving this article, the publisher recognizes that the U.S. Government retains nonexclusive, royalty-free license to publish or reproduce the published form of this contribution, or to allow others to do so, for U.S. Government purposes. Los Alamos National Laboratory requests that the publisher identify this article as work performed under the auspices of the U.S. Department of Energy. Los Alamos National Laboratory strongly supports academic freedom and a researcher's right to publish; as an institution, however, the Laboratory does not endorse the viewpoint of a publication or guarantee its technical correctness.

LOS ALAMOS
PROJECT
MAIN GATE
PASSES MUST BE
PRESENTED TO
GUARDS

NATIONAL SECURITY RESEARCH CENTER





National Security Research Center

Titan on the Red (advanced AI / ML system)



Rizwan Ali
Director, NSRC
February 2021





BACKGROUND



National Security Research Center

- Born out of a need to make LANL Weapons Program archival material available to scientists and engineers
- Material is the end result of **decades of consolidation** of mini-libraries and mini-archives at LANL
- Latest consolidation brought together LANL's **digital archives** and **physical archives**
- Houses **75+ years** of nuclear weapons research, designs, procedures, videos, photos, and other reports



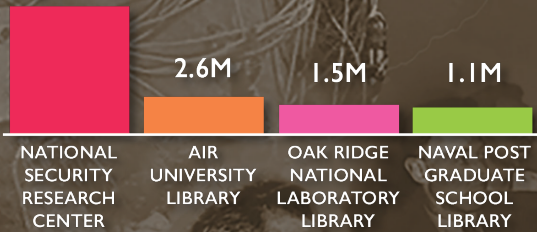
NSRC offers services similar to major university research libraries



LIBRARY COMPARISON (BY HOLDINGS)

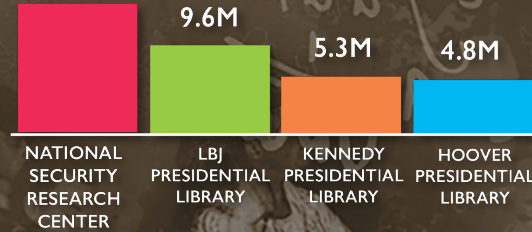
MAJOR NATIONAL SECURITY LIBRARIES

14M



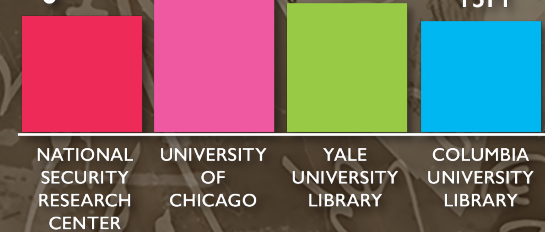
MAJOR ARCHIVAL LIBRARIES

14M



MAJOR UNIVERSITY LIBRARIES

14M

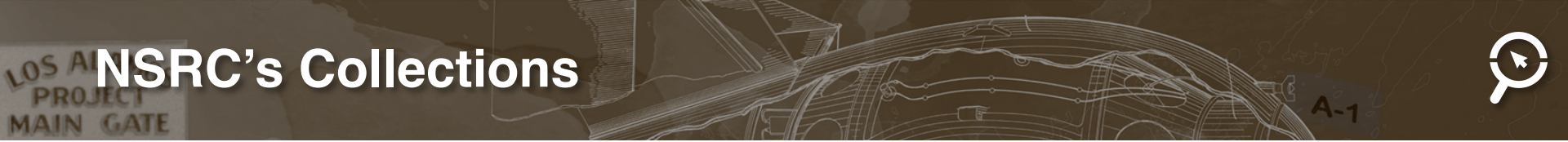


NATIONAL SECURITY
RESEARCH CENTER



THE COLLECTIONS





NSRC's Collections

- Not a traditional archive.
- NSRC's holdings are *active collections* **used daily by researchers** at LANL and throughout the NSE.
- Contains *millions* of unique and historically significant items documenting the U.S. nuclear weapons design, development, testing, certification and assessment activities.

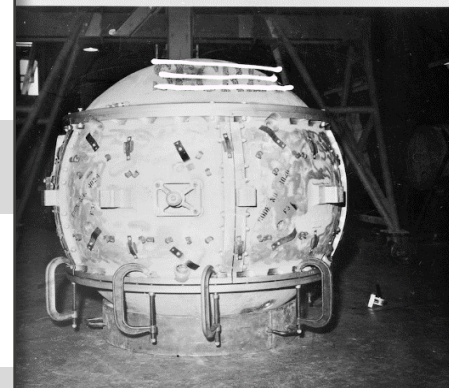
SAMPLE OF NSRC HOLDINGS
Nuclear weapons designs
Assembly drawings
Nuclear test diagnostic data
Pre-shot reports
Post-shot reports and analyses
Supporting and confirmatory experiments
Subcritical experiments
Stockpile stewardship experiments
Milestone reports
Classified research journals
US/UK collaboration documents
Internal memoranda
Working group notes
Histories
Code manuals
Test problems

Vast majority of the material is not duplicated anywhere else in the NSE

Sample of NSRC's Collections



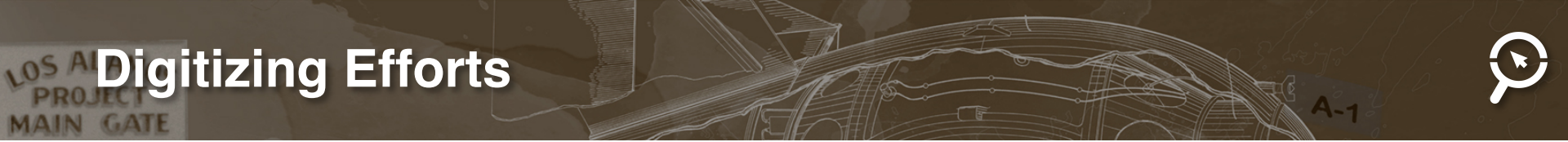
COLLECTION	CONTENT
Rocky Flats Collection	<ul style="list-style-type: none">• Detailed pit production, certification, and use information• Technical reports describing efforts by researchers to develop new pit technology
Classified Reports Collection	<ul style="list-style-type: none">• Studies of Soviet missile defense• Effects of underwater blasts on submarine integrity• Requirements for strategic stability• Development of warheads for advanced delivery systems• Special processes for pit welding
Top Secret Collection	<ul style="list-style-type: none">• From the intelligence community, Joint Chiefs of Staff, the Pentagon, and other government agencies
United Kingdom Collection	<ul style="list-style-type: none">• 7,000+ accountable docs• Originally supported the 1958 Mutual Defense Agreement• Relevant in weaponry design, naval nuclear propulsion, and nuclear threat reduction
Wellnitz Collection	<ul style="list-style-type: none">• Legacy weapons engineering drawings, hydro test records, and material• NSTEC (Nevada National Security Site) vault collection• Old J-Division (field testing) collection• New material from weapons physicists
Los Alamos Historical Collection	<ul style="list-style-type: none">• Archival material related to the Manhattan Project• Correspondences, films, reports, engineering drawings, lab notebooks



Tightening the Polar Cap to the Zone Segments.



DIGITIZING AND CATALOGING



Digitizing Efforts

- Less than 10% of the physical collection is digitized
- Decades-long project to digitize the following media types:

---> *Paper*

---> *Engineering drawings*

---> *Microfilm*

---> *Microfiche*

---> *Photos*

---> *Aperture cards*

---> *Video*

---> *Motion picture film*

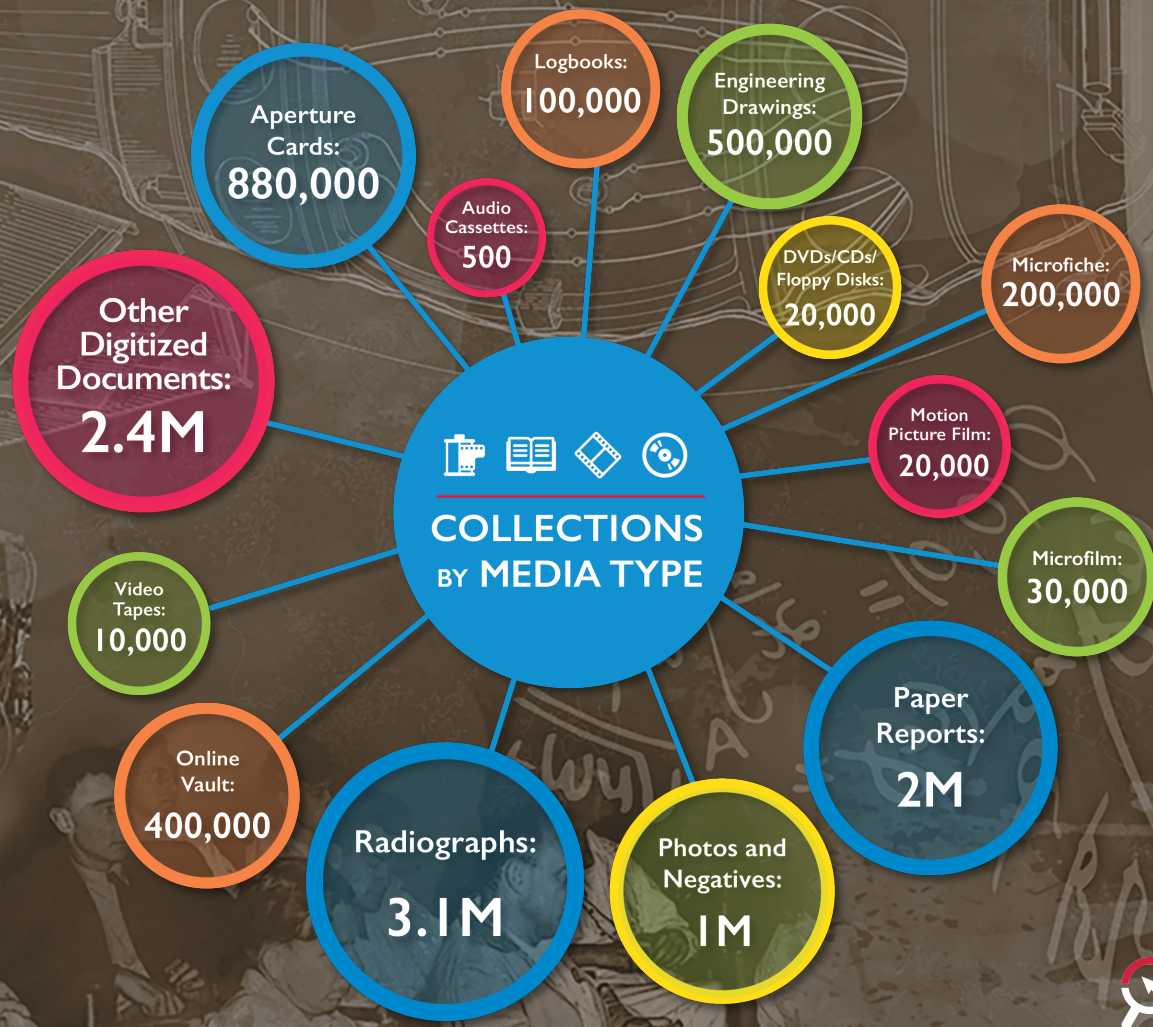
---> *Audio*

---> *Notebooks*

---> *Negatives*

---> *Card catalog*

Backlog in digitizing cannot be addressed with current processes



Digitizing Efforts – Historical Rates



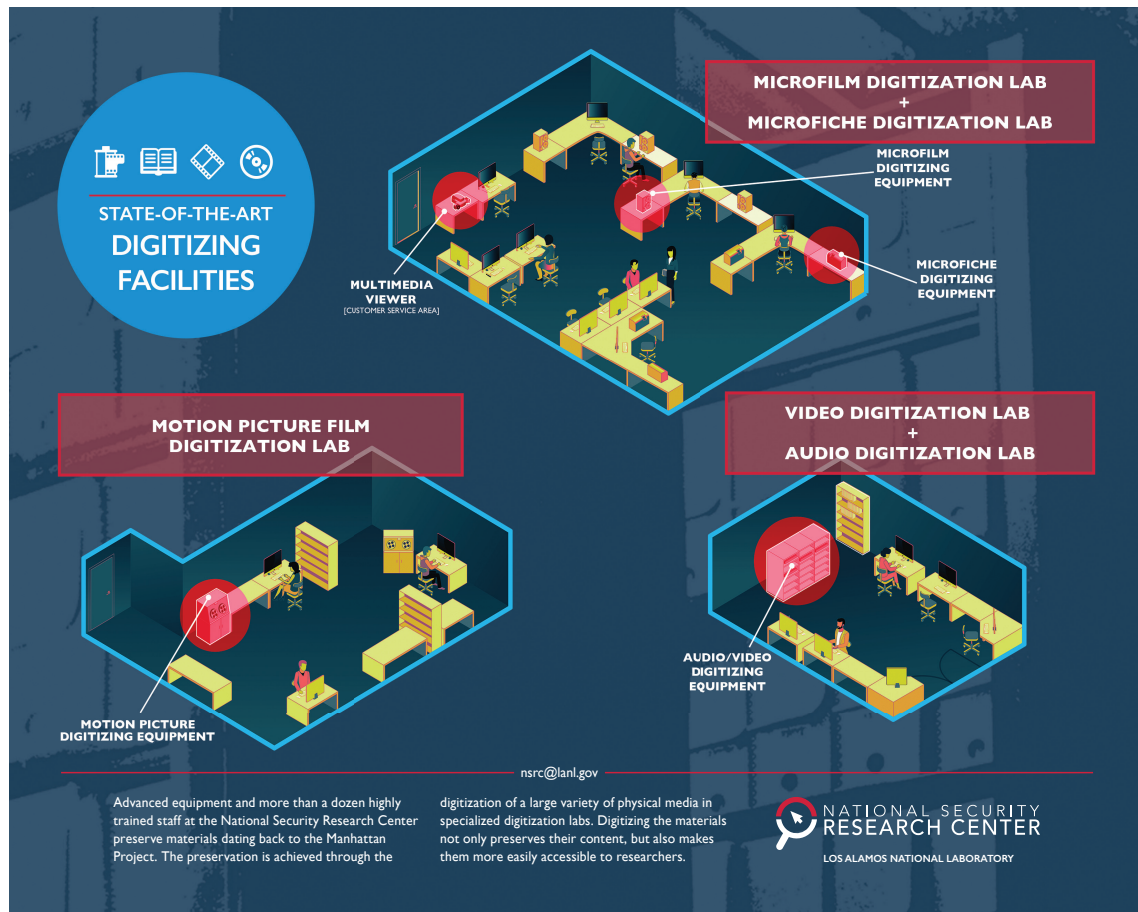
- ~1,300 videos per year
 - Total collection: ~10K (6 videos per day)
 - ~8 years to digitize
- ~15 microfilm rolls per year
 - Total collection: ~30K (1,500 – 2,000 pages per film roll)
 - ~2,000 years to digitize
- ~2,200 microfiche sheets per year
 - Total collection: ~200K (10 – 20 pages per sheet)
 - ~93 years to digitize
- ~52 "banker boxes" of paper digitized per year
 - Total collection: ~40K (~2,000 pages per box)
 - ~760 years to digitize

Digitization efforts limited by equipment, software, processes, training, facilities

Status of Digitization Services



- NSRC stood up six new digitization labs in FY20:
 - Video Tape Digitization Lab
 - Audio Tape Digitization Lab
 - Microfilm Digitization Lab
 - Microfiche Digitization Lab
 - Motion Picture Film Digitization Lab
 - Specialized Digitization Lab (paper, engineering drawings, photos, lab notebooks, aperture cards, etc.)
- Established digitization equipment certification program:
 - All applicable staff to be certified on equipment and archival standards



Refocus Attention on Highest Priority Media



- Video (VHS, Betamax, etc.)

- Fastest **deteriorating media** in our collections
- Rate of digitization may not address deteriorating media
- Increased speed through number of digitization suites / FTEs
- **Estimated completion: Down from 8 years to 5 years**

Video	Legacy	New
Quantity	10,000	
FTEs	1.0	1.5
Digitization suites	2	4
Rate per month	108	170
Years to complete	8	5

- Microfilm – Of immediate relevance to Weapons Program

- New production-scale equipment and software
- Revamped processes
- Increase of personnel from 0.1 to 2
- **Estimated completion: Down from 2,000 years to 16 years**

Microfilm	Legacy	New
Quantity	30,000	
FTEs	0.1	2
Digitization suites	1	3
Rate per month	1.3	160
Years to complete	2,000	16

- Microfiche – Of immediate relevance to Weapons Program

- New production-scale equipment and software
- Revamped processes
- Increase of personnel from 1 to 3
- **Estimated completion: Down from 93 years to 19 years**

Microfiche	Legacy	New
Quantity	200,000	
FTEs	1	3
Digitization suites	1	4
Rate per month	180	900
Years to complete	93	19



Cataloging / Indexing Efforts

- Digital Collections

- Wide variety of metadata and indices ranging from nonexistence to rudimentary to robust
- *Search functions range from nonexistence to difficult to use*

- Physical Collections

- *Poor to nonexistent metadata*
- *Limited indices*
- *Classified Reports Collection has an old-fashioned card catalog*

- Cataloging rate:

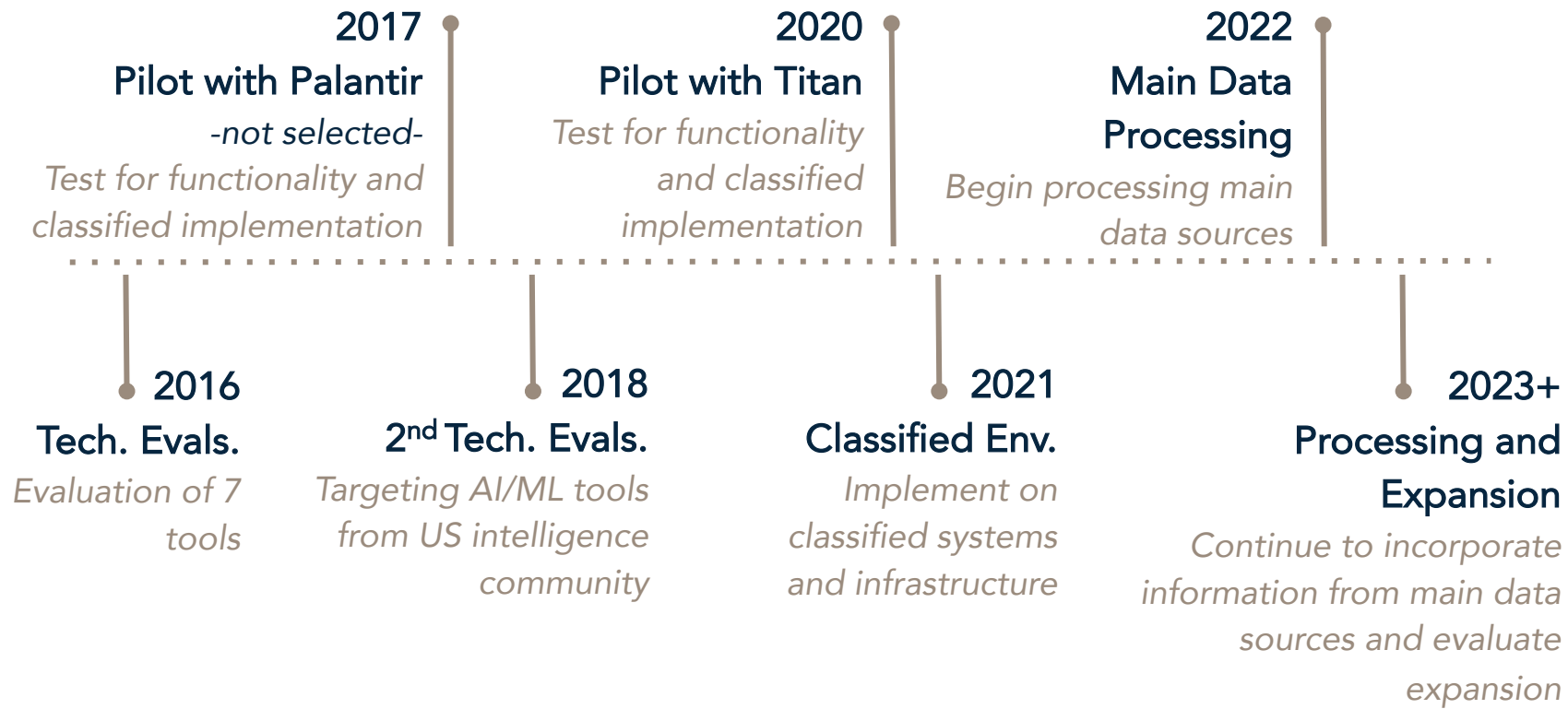
- *Currently using **fully manual** process*
- *Process takes 10-30 minutes per document*
- *Total number of **digitized files** growing rapidly*

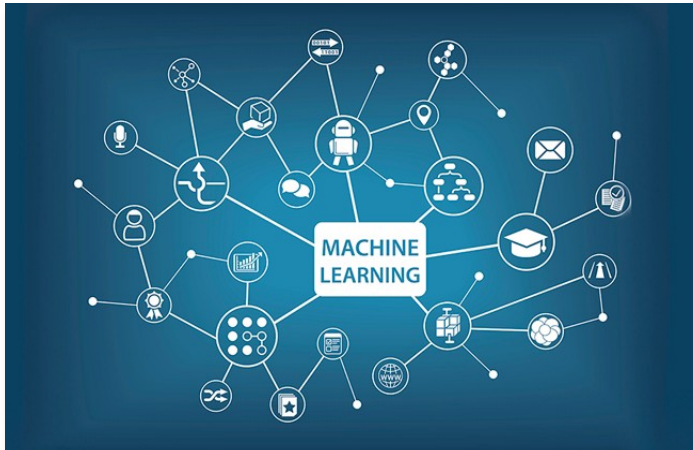
Cataloging snapshot of one digitized collection	
Quantity	2.4 million
FTEs	1.5
Rate per month	486
Years to complete	412

Backlog in cataloging/indexing cannot be met with manual approach



LANL Automated Cataloging/Search Project Timeline





- **System must be able to:**

- *Automatically extract metadata from digitized documents*
- *Perform natural-language search*
- *Enforce security classification and NTK protocols*
- *Be expandable to ingest various data stores (Online Vault, PDMLink, shared drives, SharePoint, etc.)*
- *Utilize commercial, public domain, and LANL ontologies*

- **Core Data Platform:**

- *Company found through contacts in the US Intelligence Community*
- *Company had done similar work for several of the IC's 3-letter agencies*
- *Successfully completed funded (\$800K) 6-month demo on Yellow in FY20 and passed Red Team evaluation*
- *Brought capability on-contract for Red in FY21*

- **FY21 efforts funded through combination for funds from NA-115 and LANL**

LOS ALAMOS
PROJECT
MAIN GATE
PASSES MUST BE
PRESENTED TO
GUARDS

POST
No. 1

QUESTIONS?

A-1



LOS ALAMOS
PROJECT
MAIN GATE
PASSES MUST BE
PRESENTED TO
GUARDS

NATIONAL SECURITY RESEARCH CENTER

